

# Biostatistical Issues in the Design and Analysis of Animal Carcinogenicity Experiments

by Christopher J. Portier

Two-year animal carcinogenicity experiments are used to evaluate the potential carcinogenicity from exposure to chemicals. The choice of exposure levels, the allocation of animals to doses, the length of exposure, and the choice of interim sacrifice times all affect the power of statistical tests for carcinogenic effects and the variance of interpolated estimates of carcinogenic risk. In this paper, one aspect of this problem is considered: the ability of tumor incidence data to provide information on carcinogenic mechanism and the optimal choice of design parameters with which to achieve this purpose. The direct application of biochemical data to the estimation of carcinogenic risk is also discussed in detail.

## Simple Stage Model

The mechanism by which chemicals induce carcinogenic response in test animals can be an important factor in estimating the potential carcinogenic risk resulting from human exposure. The primary method used by U.S. regulatory agencies has been to estimate cancer risks using data from 2-year animal experiments and conservative models for estimating low-dose risks. However, there has been increasing pressure on these agencies to use mechanistic models for the estimation of carcinogenic risks. Among the potential models for use, the multistage models of cancer (1-3) that include clonal expansion of cells in the various stages of carcinogenesis have received the most attention. Several authors have suggested the use of one specific form of this class of models, a simple two-stage model of carcinogenesis used extensively by Moolgavkar and co-workers (4,5) and Cohen and co-workers (6-8). Figure 1 illustrates this model. Basically, normal cells are transformed (via mutation) into premalignant or initiated cells. These initiated cells proliferate or die out via a simple birth-death process. They can also undergo a second transformation that results in a malignant cell and may eventually grow into a tumor.

The shape of the dose-response curve for carcinogenesis has a significant impact on low-dose estimates of carcinogenic risks. Models for which the slope of the dose-response curve is positive and finite at dose zero are referred to as "low-dose linear" models. For these models, small changes of dose in the low-dose range would result in proportional increases in the probability of cancer. Models for which the slope of the dose-response curve is zero or negative in the low-dose range are referred to as "nonlinear" models. For these models, a small increase in dose in the low-dose range will result in almost no change in the risk

of cancer. It has been shown that the usual animal carcinogenicity experiment provides very little information on dose-response shape and that models that yield widely divergent low-dose risks will adequately fit most data.

Portier and Edler (9) considered the ability of tumor incidence data to differentiate between various mechanisms of carcinogenesis within the context of the two-stage model (Fig. 1). Using suggestions of others (10,11), they classified carcinogenic effects into three basic classes depending upon how the dose effect is incorporated into the model. "Initiators" are defined as those carcinogens that alter the rate at which cells move from the normal state to the initiated state ( $\mu_1$  in Fig. 1). "Promoters" are thought to be chemicals that act directly on the birth rate of initiated cells ( $\beta$ ) by clonally expanding the numbers of these cells. The final mechanistic class used by Portier and Edler was labeled "completer." These are chemicals that affect the rate at which initiated cells are transformed into malignant cells, thus completing the carcinogenic process.

These mechanistic labels for carcinogenic action are basically derived from a type of carcinogenesis experiment known as the initiation-promotion-initiation (IPI) experiment. In these IPI experiments, a single dose of an initiator is given to the test animals at the start of the experiment. This is followed by chronic exposure to a promoter and, after some time, the application of another initiator. The order in which the chemicals are given is

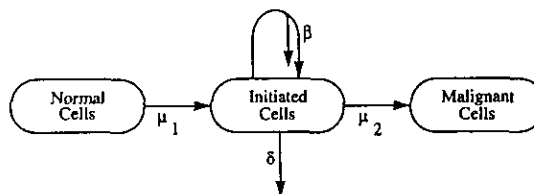


FIGURE 1. A simple two-stage model of carcinogenesis.

National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709.

This paper was presented at the International Biostatistics Conference on the Study of Toxicology, May 13-25, 1991, in Tokyo, Japan.

crucial to the rate of tumor formation. That is, if the promoter is given first, followed by the initiator, very few, if any, tumors are formed. It is thought that the initiator interacts with the DNA of normal cells causing mutations which somehow predispose these mutated cells to carcinogenesis. The promoter is thought to increase the clonal growth of only the initiated cell, allowing the numbers of these cells to increase rapidly relative to the normal cells. Finally, the second initiator (or completer) completes the carcinogenic process by causing a second mutation in the initiated cell, which results in the formation of tumors.

In an attempt to improve the estimation of low-dose risks, it has been proposed that mechanistic models of carcinogenesis be used in the risk assessment process. The advantage of mechanistic models of carcinogenesis over more empirical models is that it is believed that different carcinogenic mechanisms will result in different dose-response shapes. From the discussion above, if this is true, then information on the mechanism of action of a carcinogenic substance will result in improved low-dose risk estimates. For example, it is widely held that chemically induced mutations of the type resulting from initiators and completers are low-dose linear (10,11). The mechanisms that lead to chemically induced promotion are thought to be nonlinear. These theories are highly speculative (12); yet, we can look at the operating characteristics of applying them to see if further research into their use is warranted on statistical grounds.

In their analysis, Portier and Edler (9) were able to show that the usual design of the animal carcinogenesis experiment provided little information that could be used to differentiate between linear initiation/completion effects and nonlinear promotion effects. Their basic approach was as follows. Control rates for the baseline (untreated) parameters in the two-stage model were chosen using historical information on a large population of control animals (13). Various levels of dose effects (low tumor yield to high tumor yield) were determined for each set of parameters for each potential mechanism (initiation, promotion, and completion). Given one such hypothesized model, they then simulated the results of an animal carcinogenicity experiment based on a particular choice of design parameters. These design parameters included the standard design (no interim sacrifices, three dosed groups and a control group, 50 animals per group, doses in the relative magnitude of 0, 0.25, 0.5 and 1) and eight other designs that added start/stop dosing and interim sacrifices. For each simulated data set, parameters for four two-stage models (the four two-stage models fit to the data included a model where the effect of dose was an initiation effect, a model for which the effect of dose was a completion effect, a model for which the effect of dose was a promotion effect, and a global model which allowed for all three types of effects) were estimated based on maximum likelihood estimation. Likelihood ratio techniques were used to determine how often each of the three singular-effect models (initiation only, promotion only, and completion only) described the data as well as the model that allowed for all three effects. Under this modeling scheme, they were then able to study the effect of changing the design of the carcinogenesis experiment on the rate of rejection of the various models.

Under the usual design of animal carcinogenicity experiments, it was generally found that all of the models fit the data well, regardless of the underlying model. When data were generated assuming a linear initiation-only effect, the initiation model and

the completion model provided as good a fit as the global model in virtually all cases (>99% of the time). All three models were accepted in 85–95% of the cases studied (depending on the magnitude of the assumed dose effect). When a linear completion effect was assumed, similar results were obtained. If the completion effect was assumed to be a function of dose squared (quadratic completion model), approximately 98% of the cases were adequately fit by either a linear initiation model or a quadratic completion model and 74–93% of the cases studied were fit by all three models.

For promotion effects, Portier and Edler (13) considered dose effects on the birth rate of initiated cells that were functions of dose raised to the first power (linear promoter model) up through the fourth power (quartic promoter model), resulting in four basic models. When the assumed model was based upon a linear promotion effect, all three models (linear initiator, linear completer, and linear promoter) fit the data in approximately 95% of the cases studied. As the shape of the dose effect became more nonlinear, it was possible to reject the linear initiation model and the linear completion model with greater power. For a quartic promotion model, the initiation and completion models could be rejected in 15–50% of the cases studied, whereas the promotion model was accepted in about 95% of the cases. Thus, only for a highly nonlinear promotion effect was there any strong degree of differentiation between these models when the usual design of the long-term animal carcinogenicity experiment was used.

Based on these results, Portier and Edler (13) considered several alternative designs. When the underlying model was based on a linear initiation effect, there was a slight improvement in the probability of rejecting the promotion model; going from 4 to 12% in the usual design to 5 to 26% in the start-stop designs. The ability to reject the completion model did not change noticeably when the start-stop designs were used. When the underlying model was a linear completion model, the results were similar to the initiation model. For an underlying quadratic completion model, it was more difficult to reject the promotion model and easier to reject the initiation model. However, these differences were small. Finally, for underlying promotion models, the use of start studies reduced the percentage of times we could reject the initiation model or the completion model when compared to the usual bioassay design. This is due to the fact that fewer doses were used (in favor of equal doses over varying time spans) and that both magnitude of dose and length of dosing play an important role in differentiating between initiators/completers and promoters.

The results of this research suggest that differentiation between initiation and completion is best accomplished with start-stop dosing experiments; that the rejection of a promotion model when the dose effect in the underlying model is an initiation effect or a completion effect is improved by using an early exposure-stop exposure group in the experiment; and that if the underlying model is a nonlinear promotion model, it is better to use multiple doses than start-stop dosing at similar levels of exposure. Finally, it was found that all the designs were generally poor for distinguishing mechanism.

## Damage-Fixation Multistage Model

After reviewing the results of these experiments, it was clear that tumor incidence data could not be reliably used to determine

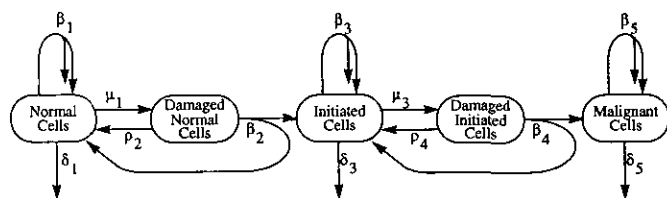


FIGURE 2. A two-stage model of carcinogenesis (damage-fixation multistage model) that incorporates damage and repair.

how treatment affects mutation rates and birth/death rates in multistage models of carcinogenesis. One way out of this dilemma is to use other toxicological data such as the size and number of initiated cells (14). However, current research is focusing on the use of direct mechanistic information on the carcinogenicity of a compound to estimate the tumor incidence rate and then to use the tumorigenesis data from the animal carcinogenicity experiment to validate the model. To be able to do this, a slightly different model of carcinogenesis is needed. It has been noted that a mutation is itself the result of a process that involves at least two steps. In the first step, damage must occur. For a mutation to occur, this damage must then be fixed by replication of the damaged cell. It is also clear that this damage may not persist forever but may be repaired via numerous mechanisms in the cell. Thus, several events are competing or combining to result in a single mutation. The model presented in Figure 1 does not explicitly account for this more detailed mutation process.

The model of carcinogenesis illustrated in Figure 2 is also a two-stage model of carcinogenesis with clonal expansion of all cell types. This model is referred to (15,16) as the damage-fixation multistage model (DFM). The model has five cell types: normal cells, two types of damaged cells and two types of mutated cells in which the DNA damage has been fixed by cell replication. For a cell to become malignant, it must pass from the normal state through each of the mutational states. The dynamics of the model can be illustrated as follows. Normal cells are allowed to divide and die or differentiate. Normal cells transform into damaged cells via some type of genetic aberration (e.g., formation of DNA adducts, single-strand breaks, chromosomal translocation). The genetic aberrations in these damaged cells are assumed to pertain to a single strand and can be repaired, returning the cell to its normal state. When cell division occurs in these unrepaired cells, the DNA damage is fixed in one of the daughter cells resulting in the creation of a single mutated cell. The other daughter cell is derived from the strand of DNA without damage and is thus a normal cell. The process of damage, repair, birth, and death is repeated in the second stage.

The DFM model allows for the direct inclusion of biochemical data into an analysis of carcinogenic mechanism and the estimation of carcinogenic risk. These data include the rate of formation of DNA adducts (DNA damage), the rate of DNA repair, and the rate of cell replication and death. First consider the rate of DNA damage. For example, it has been demonstrated that administration of 3 nmole of 7,12-dimethylbenz[*a*]anthracene (DMBA) to Swiss mice results in a binding of 1 nmole of DMBA per mole of DNA-P 24 hr after exposure. If this damage is critical to the conversion of normal cells into first stage cells, then these

data can be incorporated into the model directly by setting  $\mu_0$  equal to the rate of binding per unit time (being certain to express this rate in the proper units of rate of damage per cell per unit of time). However, it is more likely that some specific type of damage is inducing the mutation. In this case, the relative change in the nonspecific damage as a function of dose can be used as a surrogate for the relative change in the specific binding. In the example above, a dose of DMBA of 150 nmole resulted in a binding of 14 nmole of DMBA per mole of DNA-P 24 hr after exposure (i.e., a 50-fold increase in dose of DMBA resulted in a 14-fold increase in binding). Thus, even though the specific adduct that induces the mutation is unknown, a 14-fold increase in the specific adduct which induces the mutation could be assumed when going from a dose of 3 nmole to 150 nmole of DMBA.

Estimates of the repair rates for DNA damage can be obtained in similar ways. The most obvious method is to directly measure the activity of proteins involved in the DNA repair process such as O<sup>6</sup>-alkylguanine-DNA alkyltransferase. However, like nonspecific versus specific DNA damage, the specific repair mechanism is generally unknown, and so the proteins involved in its repair are also unknown. One way to avoid this would be to obtain the DNA damage rate from biochemical experiments using simple compartment models and estimate the repair rate from data on tumor incidence or data on the size distribution of cells in each stage. This approach is likely to lead to statistical dependencies in the estimated parameters and large uncertainty in the estimated tumor damage at several different time points following exposure. In this case, the differences over time in the amount of DNA damage should yield an estimate of the repair rate. This method is preferable to using tumor incidence data or cell count data because the estimate of the repair rate would come directly from data on DNA damage and would not be dependent on the applicability of the model. A third method for estimating DNA repair would be to see how much of the damage could be fixed at different times following exposure to the compound. For example, in the two-stage experimental protocol described above, waiting varying lengths of times from initiation to the start of promotion allows for a longer period of DNA repair and the level of DNA repair can be estimated.

There are also a variety of ways in which cell replication rates can be measured in animal tissues. These methods are very direct in the sense that, in a fixed period of time, they label all cells that have undergone replication. These techniques can even be used with other cellular techniques such as staining for enzyme alteration. In this case, it is possible to directly measure the rate of cell replication in normal cells, in initiated cells (provided a probe exists for staining the cells or in some other way labeling them), and in malignant cells. The technology for the direct estimation of cell death/differentiation rates are currently being developed. As these become available, they can be directly incorporated into the DFM model. Until then, information on the size distribution and number of initiated calls can be used to estimate this parameter.

The approach of estimating these parameters from data other than tumor incidence data is illustrated in Portier and Kopp-Schneider (15) and Kopp-Schneider et al. (16).

## Discussion

This paper has reviewed some of the problems concerning the characterization of mechanistic models of carcinogenesis using

tumorigenesis data. On strictly statistical terms, it was shown that the usual two-year rodent carcinogenicity experiment does not provide sufficient information to be able to differentiate between some basic mechanistic models of carcinogenesis. Modification of the bioassay design to include time-varying doses did not dramatically improve this problem.

A two-stage model of carcinogenesis that allows for the direct inclusion of biochemical data into the estimation of carcinogenic risks was reviewed. This approach has the advantage that the tumor incidence data from the long-term animal carcinogenesis experiment and/or the cell-kinetic information on cells in the different stages can be used to validate the model. However, there are numerous problems with the use of this modeling approach for risk estimation. It is imperative that an attempt is made to validate the model predictions using the available toxicological data (e.g., tumor incidence data from carcinogenicity experiments, papilloma counts from skin painting studies, etc.). This validation needs to be done with extreme caution because goodness-of-fit tests that would be used in this context are generally insensitive to moderate changes in the model parameters or even slightly different models. Not only are there statistical problems with this approach, but there are inadequacies in the biological description of DNA damage and repair in the DFM model. The DNA of a cell can be damaged in many places in many different ways; it may be that cells with multiple DNA damage are more (or less) susceptible to replication and/or mutation than are cells with little DNA damage. The model presented here assumes that DNA damage is either present or not, thus using only partial information concerning the process. The rate of DNA repair in any one cell is likely to be tied to the amount of damage in that one cell, a concept that is not allowed in the current model formulation. Other issues such as strand-specific DNA repair, preferential DNA repair, and DNA hot spots will also limit the usefulness of models of this type. Cell replication rates must also be applied cautiously; if the increased cell replication only pertains to a small fraction of the total tissue, this must be accounted for. Finally, all of these rates may change with age as well as dose, thus the experiments in which these biochemical parameters are obtained must include several ages as well as several doses.

## REFERENCES

1. Armitage, P., and Doll, R. The age distribution of cancer and a multistage theory of cancer. *Br. J. Cancer* 8: 1-12 (1954).
2. Neyman, J. and Scott, E. Statistical aspects of the problem of carcinogenesis. In: *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, CA, 1967, pp. 745-776.
3. Whittemore, A., and Keller, J. Quantitative theories of carcinogenesis. *SIAM Rev.* 20: 1-20 (1978).
4. Moolgavkar S., and Venzon, D. Two-event models for carcinogenesis: incidence curves for childhood and adult tumors. *Math. Biosci.* 47:55-77 (1979).
5. Moolgavkar, S., and Knudson, A. Mutation and cancer: a model for human carcinogenesis. *J. Nat. Cancer Inst.* 66:1037-1052 (1981).
6. Cohen, S., and Ellwein, L. Cell proliferation in carcinogenesis. *Science* 249: 1007-1011 (1990).
7. Ellwein, L., and Cohen, S. A cellular dynamics model of experimental bladder cancer: analysis of the effect of sodium saccharin in the rat. *Risk Anal.* 8: 215-221 (1988).
8. Greenfield, R., Ellwein, L., and Cohen, S. A general probabilistic model of carcinogenesis: analysis of experimental urinary bladder cancer. *Carcinogenesis* 5:437-445 (1984).
9. Portier, C., and Elder, L. Two-stage models of carcinogenesis, classification of agents and design of experiments. *Fundam. Appl. Toxicol.* 14: 444-460 (1990).
10. Thorsland, T., Brown, C., and Charnley, G. Biologically motivated cancer risk models. *Risk Anal.* 7: 109-119 (1987).
11. Moolgavkar, S. Model for human carcinogenesis: action of environmental agents. *Environ. Health Perspect.* 50: 285-291 (1983).
12. Barrett, J. C., and Wiseman, R. Cellular and molecular mechanisms of multistep carcinogenesis: relevance to carcinogen risk assessment. *Environ. Health Perspect.* 76: 65-70 (1987).
13. Portier, C., and Bailer, A. Two-stage models of carcinogenesis for historical control animals from the National Toxicology Program. *J. Toxicol. Environ. Health* 27: 21-45 (1989).
14. Moolgavkar, S., Luebeck, E. G., de Gunst, M., Port, R. and Schwartz, M. Quantitative analysis of enzyme-altered foci in rat hepatocarcinogenesis experiments. I: Single agent regimen. *Carcinogenesis* 11: 1271-1278 (1990).
15. Portier, C., and Kopp-Schneider, A. A multistage model of carcinogenesis incorporating DNA damage and repair. *Risk Anal.* 11: 535-543 (1991).
16. Kopp-Schneider, A., Portier, C. J., and Rippman, F. The number and size of detectable clones of initiated cells in a damage-fixation model in initiation/promotion experiments. *Math. Biosci.* 105: 139-166 (1991).
17. Portier, C., and Kaplan, N. The variability of safe dose estimates when using complicated models of the carcinogenic process. a case study: methylene chloride. *Fundam. Appl. Toxicol.* 13: 533-544 (1989).